What is Claimed is:

1. A sequence based indexing and retrieval method for text documents, comprising the steps of:

(a) generating a query token sequence, having at least a query token, from a query submitted by a user;

(b) generating at least a representative token sequence, having at least a document token, from each of said text documents that contain at least one token of said query token sequence;

(c) measuring a similarity between each of said representative token sequences and said query token sequence by:

(c.1) determining a token appearance score by measuring a token appearance of said representative token sequence with respect to said query token sequence;

(c.2) determining a token order score by measuring a token order of said representative token sequence with respect to said query token sequence; and

(c.3) determining a token consecutiveness score by measuring a token consecutiveness of said representative token sequence with respect to said query token sequence; and

(d) retrieving said text documents in responsive to said similarity of said representative token sequence with respect to said query token sequence with a ranking order in accordance with said token appearance score, said token order score, and said token consecutiveness score, provided that for a document with two representative token sequences, its similarity is determined by the representative token sequence with a higher score.

2. The method, as recited in claim 1, wherein the step (c.1) comprises the sub-steps of:

(c.1.1) consulting an index of said text documents to determine the weight of each token in said query token sequence;

(c.1.2) calculating a sum of the weights of the query tokens that appear in said representative token sequence; and

(c.1.3) outputting said token appearance score of said token appearance by calculating a fraction of said sum divided by the total weight of all query tokens.

3. The method, as recited in claim 2, wherein said weight of said query token in said query token sequence is measured by determining a token frequency of said query token in said text documents.

4. The method, as recited in claim 1, wherein the step (c.2) comprises the sub-steps of:

(c.2.1) determining a length of the longest common subsequence of said representative token sequence and said query token sequence;

(c.2.2) determining a length of said representative token sequence;

(c.2.3) determining a length of said query token sequence; and

(c.2.4) outputting said token order score of said token order by calculating a fraction of said length of said longest common subsequence divided by an average sum of said length of said representative token sequence and said length of said query token sequence.

5. The method, as recited in claim 3, wherein the step (c.2) comprises the sub-steps of:

(c.2.1) determining a length of the longest common subsequence of said representative token sequence and said query token sequence;

(c.2.2) determining a length of said representative token sequence;

(c.2.3) determining a length of said query token sequence; and

(c.2.4) outputting said token order score of said token order by calculating a fraction of said length of said longest common subsequence divided by an average sum of said length of said representative token sequence and said length of said query token sequence.

6. The method, as recited in claim 1, wherein the step (c.3) comprises the sub-steps of:

(c.3.1) determining a relative distance between a positional differentiation of each adjacent document tokens and a positional differentiation of said adjacent document tokens in said query token sequence; and

(c.3.2) outputting said token consecutiveness score of said token consecutiveness by calculating a fraction of a sum of the inverses of said relative distances divided by the number of pairs of adjacent tokens, which equals the length of said representative token sequence less one.

7. The method, as recited in claim 3, wherein the step (c.3) comprises the sub-steps of:

(c.3.1) determining a relative distance between a positional differentiation of each adjacent document tokens and a positional differentiation of said adjacent document tokens in said query token sequence; and

(c.3.2) outputting said token consecutiveness score of said token consecutiveness by calculating a fraction of a sum of the inverses of said relative distances divided by the number of pairs of adjacent tokens, which equals the length of said representative token sequence less one.

8. The method, as recited in claim 5, wherein the step (c.3) comprises the sub-steps of:

(c.3.1) determining a relative distance between a positional differentiation of each adjacent document tokens and a positional differentiation of said adjacent document tokens in said query token sequence; and

(c.3.2) outputting said token consecutiveness score of said token consecutiveness by calculating a sum of the inverses of said relative distances with respect to said representative token sequence.

9. The method, as recited in claim 8, wherein said similarity of said representative token sequence is calculated with respect to said query token sequence by summing said token appearance score, said token order score, and said token consecutiveness score, wherein said ranking order of said text documents is determined by a weighted sum of said token appearance score, said token order score, and said token consecutiveness score of each of said representative token sequences of said text documents.

10. The method as recited in claim 1, in step (b), further comprising a step of selecting at least a candidate document from said text documents, wherein one of said text documents is selected to be said candidate document when said text document contains at least one token of said query token sequence.

11. The method as recited in claim 9, in step (b), further comprising a step of selecting at least a candidate document from said text documents, wherein one of said text documents is selected to be said candidate document when said text document contains at least one token of said query token sequence.

12. The method as recited in claim 10, in step (b), further comprising a step of consulting an index of said text documents to establish said candidate document, wherein tokens that also appear in the query token sequence are collected to form a document token sequence for each document and the two longest segments of said document token sequence are selected as representative token sequences wherein the positional differentiation of each adjacent document tokens is no larger than a predetermined positioning value while said corresponding text document is selected as the said candidate document.

13. The method as recited in claim 11, in step (b), further comprising a step of consulting an index of said text documents to establish said candidate document, wherein tokens that also appear in the query token sequence are collected to form a document token sequence for each document and the two longest segments of said document token sequence are selected as representative token sequences wherein the positional differentiation of each adjacent document tokens is no larger than a predetermined positioning value while said corresponding text document is selected as the said candidate document.

14. The method as recited in claim 10, in step (b), further comprising a step of retaining said candidate document to be used for measuring said similarity with respect to said query token sequence, wherein the said candidate document is retained when said candidate document contains a token that has a weight no less than a predetermined fraction of the total weight of query tokens.

15. The method as recited in claim 11, in step (b), further comprising a step of retaining said candidate document to be used for measuring said similarity with respect to said query token sequence, wherein the said candidate document is retained when said candidate document contains a token that has a weight no less than a predetermined fraction of the total weight of query tokens.

16. The method as recited in claim 13, in step (b), further comprising a step of retaining said candidate document to be used for measuring said similarity with respect to said query token sequence, wherein the said candidate document is retained when said candidate document contains a token that has a weight no less than a predetermined fraction of the total weight of query tokens.

17. The method, as recited in claim 1, wherein said text document contains Chinese characters, English words, numbers, punctuations, and symbols as said document tokens.

18. The method, as recited in claim 9, wherein said text document contains Chinese characters, English words, numbers, punctuations, and symbols as said document tokens.

19. The method, as recited in claim 13, wherein said text document contains Chinese characters, English words, numbers, punctuations, and symbols as said document tokens.

20. The method, as recited in claim 16, wherein said text document contains Chinese characters, English words, numbers, punctuations, and symbols as said document tokens.